

KOMPARASI ALGORITMA K-NEAREST NEIGHBOR DAN NAIVE BAIYES UNTUK MENDETEKSI DINI RESIKO KANKER SERVIKS PADA REMAJA

COMPARISON OF K-NEAREST NEIGHBOR AND NAIVE BAIYES ALGORITHM TO DETERMINE EARLY RISK OF CERVICAL CANCER IN ADOLESCENTS

Mayang Sari ^{1*}, Yusri Ikhwan ²

Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, Fakultas Teknologi Informatika, Jl. Adhyaksa No.2 Kayu Tangi Banjarmasin 70123

*E-mail: mayangsari.uniska@gmail.com

ABSTRAK

Kanker saat ini menjadi penyakit paling ditakuti oleh setiap orang. Ada banyak jenis kanker yang dapat menjadi pembunuh bagi manusia, salah satunya adalah kanker serviks. kanker serviks memiliki angka tertinggi di antara kanker lain yang paling mematikan di Indonesia. Kanker serviks sendiri terjadi karena adanya penularan virus HPV (human papillomavirus), namun pada dasarnya kanker serviks dapat dicegah sejak dini dengan memberikan pengetahuan kepada masyarakat khususnya remaja sehingga dapat menekan angka kejadian kanker serviks tersebut.

Deteksi dini sangatlah diperlukan untuk mencegah terjadinya kanker serviks, ada beberapa cara yang dapat dilakukan untuk mendeteksi dini resiko kanker serviks pada remaja salah satunya dengan memanfaatkan media teknologi. Dalam penelitian ini metode yang digunakan untuk memprediksi resiko kanker serviks pada remaja dilakukan dengan mengkomparasi dua algoritma yaitu K-Nearest Neighbor dan Naïve Bayes yang mana masing-masing algoritma diuji untuk menghasilkan prediksi dengan tingkat akurasi yang maksimal.

Berdasarkan hasil implementasi dan pengukuran kedua algoritma, diperoleh algoritma terbaik yaitu Naïve Bayes yang mampu memprediksi dengan tingkat akurasi 85.71 sedangkan K-Nearest Neighbor mampu memprediksi dengan tingkat akurasi sebesar 80.95%. tetapi akurasi ini masih belum bisa dinilai excellent (sangat baik) untuk itu sangat perlu dilakukan pengembangan analisis dan hasil lebih lanjut, khususnya memperdalam faktor resiko kanker serviks pada remaja sehingga didapatkan hasil akurasi yang lebih optimal.

Kata Kunci : Kanker Serviks, Prediksi, K-Nearest Neighbor, Naïve Bayes

ABSTRACT

Cancer is now the most feared disease for everyone. There are many types of cancer that can be a killer for humans, one of which is cervical cancer. Cervical cancer has the highest number among the other deadliest cancers in Indonesia. Cervical cancer itself occurs because of the transmission of the HPV virus (human papillomavirus), but basically cervical cancer can be prevented early on by providing knowledge to the public, especially adolescents, so that it can reduce the incidence of cervical cancer.

Early detection is needed to prevent cervical cancer, there are several ways that can be done to detect early risk of cervical cancer in adolescents, one of them is by utilizing media technology. In this study the method used to predict the risk of cervical cancer in adolescents is done by comparing the two algorithms, K-Nearest Neighbor and Naïve Bayes, where each algorithm is tested to produce predictions with maximum accuracy.

Based on the results of the implementation and measurement of the two algorithms, the best algorithm is Naïve Bayes which is able to predict with an accuracy rate of 85.71 while K-Nearest Neighbor is able to predict with an accuracy of 80.95%. but this accuracy still cannot be considered excellent (very good) for that it is very necessary to develop further analysis and results, especially deepening the risk factors for cervical cancer in adolescents so that the results of more optimal accuracy can be obtained.

Keywords: Cervical Cancer, Prediction, K-Nearest Neighbor, Naïve Bayes

PENDAHULUAN

Teknologi saat ini sudah semakin berkembang sangat pesat, dengan merambahnya teknologi di berbagai sektor termasuk di bidang kesehatan. Dukungan teknologi di bidang kesehatan memberikan sumbangan yang besar untuk kemajuan ilmu kesehatan. Tak hanya sampai di situ, teknologi juga sangat membantu dalam memberikan pencegahan dini berbagai penyakit, salah satunya penyakit kanker.

Kanker saat ini menjadi penyakit paling ditakuti oleh setiap orang. Ada banyak jenis kanker yang dapat menjadi pembunuh bagi manusia, salah satunya adalah kanker serviks. Di Indonesia, pada tahun 2013 tercatat 98.692 estimasi jumlah penderita kanker serviks (Kementerian Kesehatan RI, 2015).

Kanker serviks memiliki angka tertinggi di antara kanker lain yang paling mematikan di Indonesia. Kanker serviks sendiri terjadi karena adanya penularan virus HPV (human papilomavirus), namun pada dasarnya kanker serviks dapat dicegah sejak dini dengan memberikan pengetahuan kepada masyarakat khususnya remaja sehingga dapat menekan angka kejadian kanker serviks tersebut.

Deteksi dini sangatlah diperlukan untuk mencegah terjadinya kanker serviks, ada beberapa cara yang dapat dilakukan untuk mendeteksi dini resiko kanker serviks pada remaja salah satunya dengan memanfaatkan media teknologi. Data mining adalah salah satu ilmu pengetahuan yang dapat mendukung proses deteksi dini resiko kanker serviks, dengan memanfaatkan data yang diperoleh dari hasil lapangan gejala-gejala kanker serviks dapat diprediksi dengan memanfaatkan algoritma data mining salah satunya algoritma K-Nearest Neighbor dan Naive Bayes yang mana masing-masing memiliki keunggulan masing-masing dalam hal klasifikasi. Kedua algoritma ini nantinya akan digunakan pada data hasil pengolahan kuisioner yang diisi oleh remaja-remaja yang mana diharapkan dapat menghasilkan tingkat akurasi yang lebih baik untuk mendeteksi dini resiko kanker serviks pada remaja.

METODOLOGI

1) Metode Pengumpulan Data

- Pengumpulan data primer, dengan melakukan kunjungan ke SMA yang telah dipilih secara acak dan membagikan angket yang harus diisi oleh

beberapa perwakilan siswa sebagai sampel.

- Pengumpulan data sekunder, diperoleh dengan menjalin kerjasama dengan Yayasan Kanker Indonesia untuk memperoleh informasi tentang gejala kanker serviks, serta dengan cara mempelajari dan mengutip dari berbagai literatur

1) Studi Kepustakaan

Studi kepustakaan dilakukan dengan mencari referensi dari beberapa buku atau jurnal penelitian yang berhubungan dengan Algoritma K-Nearest Neighbor dan Naïve Bayes serta tentang gejala-gejala kanker serviks pada remaja.

2) Wawancara

Wawancara dilakukan di Yayasan Kanker Indonesia (YKI) untuk memperoleh sumber data gejala-gejala kanker serviks pada remaja.

HASIL DAN PEMBAHASAN

Dalam penelitian ini, tahapan pengambilan dibagi menjadi dua bagian yaitu data training dan data testing. Data training Data training digunakan algoritma untuk membentuk sebuah model classifier. Data training didapat dari hasil wawancara yang dilakukan di Yayasan Kanker Indonesia sehingga didapatkan data training seperti berikut :

Row No.	RESIKO	USIA	MAKANAN I.	OLAHRAGA	MENIKAH	MEROKOK	SEKSUAL >...	KELUARGA
1	BERESIKO	USIA < 21	YA	TIDAK	TIDAK	YA	YA	TIDAK
2	BERESIKO	USIA < 21	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA
3	BERESIKO	USIA < 21	YA	TIDAK	YA	TIDAK	YA	TIDAK
4	TIDAK	USIA < 21	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK
5	TIDAK	USIA < 17	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK
6	BERESIKO	USIA < 17	YA	YA	TIDAK	TIDAK	YA	TIDAK
7	BERESIKO	USIA < 15	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA
8	TIDAK	USIA < 15	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK
9	BERESIKO	USIA < 15	TIDAK	YA	TIDAK	TIDAK	YA	TIDAK

Gambar 1 Data Training

Data berikutnya adalah data testing. Data testing digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar. Data testing diambil dari hasil kuisioner yang dilakukan di SMKN 1 Banjarmasin sebanyak 90 responden. Data kemudian diolah sehingga menghasilkan data sebagai berikut :

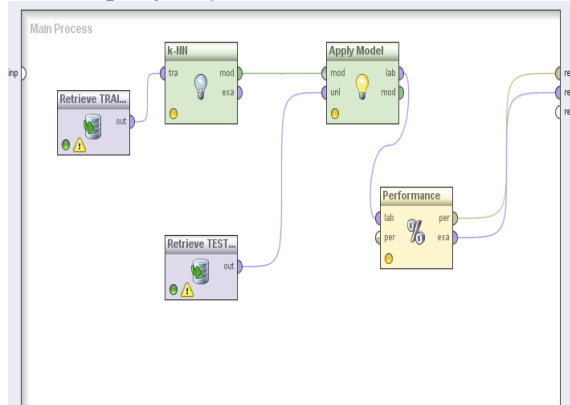
USIA	MAKANAN INSTANT	OLAHRAGA	MENIKAH	MEROKOK	SEKSUAL > 1 ORANG	KELUARGA	RESIKO
USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 17	YA	YA	TIDAK	TIDAK	TIDAK	YA	BERESIKO
USIA < 21	YA	YA	TIDAK	TIDAK	YA	TIDAK	BERESIKO
USIA < 17	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK
USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 17	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 17	YA	YA	TIDAK	YA	YA	TIDAK	BERESIKO
USIA < 17	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 21	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	BERESIKO
USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
USIA < 21	YA	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK

Gambar 2 Contoh Data Testing

Kedua data tersebut kemudian diimplementasikan untuk dilakukan pengujian. Pengujian dilakukan dengan membandingkan hasil akurasi dari dua algoritma yaitu K-Nearest Neighbor dan Naive Baiyes yang mana masing-masing algoritma memiliki keunggulan masing-masing dalam hal klasifikasi.

2) Pengujian Data Menggunakan Algoritma K-NN

Untuk menentukan apakah data tersebut masuk dalam klasifikasi beresiko atau tidak langkah pertama yang dilakukan adalah memasukkan data training ke dalam algoritma k-Nearest Neighbor. Selanjutnya data testing dimasukkan sehingga algoritma K-NN dapat menguji data tersebut berdasarkan data yang sudah dipelajarinya.



Gambar 3 Pengujian data menggunakan algoritma K-NN

Adapun hasil dari pengujian Algoritma K-Nearest Neighbor tersebut adalah sebagai berikut :

accuracy: 80.95%			
	true TIDAK	true BERESIKO	class precision
pred. TIDAK	65	0	100.00%
pred. BERESIKO	20	20	50.00%
class recall	76.47%	100.00%	

Gambar 4 Hasil Pengujian menggunakan Algoritma K-NN

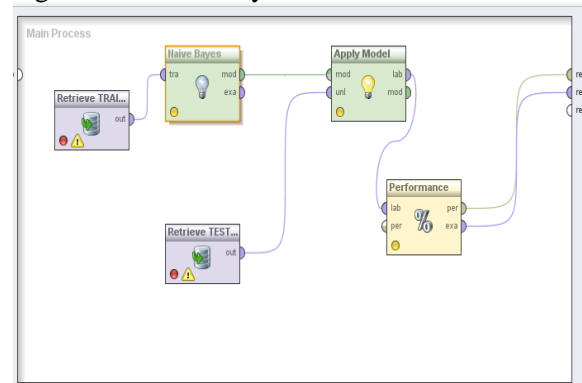
Dari hasil pengujian yang dilakukan oleh algoritma K-Nearest Neighbor didapatkan tingkat akurasi sebesar 80.95% dengan hasil prediksi menggunakan K-Nearest Neighbor sebagai berikut :

Row No.	RESIKO	confidence_	confidence_	prediction_	USIA	MAKANAN INSTANT	OLAHRAGA	MENIKAH	MEROKOK	SEKSUAL > 1 ORANG	KELUARGA	RESIKO
1	TIDAK	0	1	TIDAK	USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
2	BERESIKO	1	0	BERESIKO	USIA < 17	YA	YA	TIDAK	TIDAK	TIDAK	YA	TIDAK
3	BERESIKO	1	0	BERESIKO	USIA < 21	YA	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK
4	TIDAK	0	1	TIDAK	USIA < 17	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK
5	TIDAK	0	1	TIDAK	USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
6	TIDAK	0	1	TIDAK	USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
7	TIDAK	0	1	TIDAK	USIA < 17	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
8	BERESIKO	1	0	BERESIKO	USIA < 17	YA	YA	TIDAK	YA	YA	TIDAK	TIDAK
9	TIDAK	0	1	TIDAK	USIA < 17	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
10	TIDAK	0	1	TIDAK	USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
11	TIDAK	1	0	BERESIKO	USIA < 17	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
12	TIDAK	0	1	TIDAK	USIA < 17	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
13	TIDAK	0	1	TIDAK	USIA < 17	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK
14	TIDAK	0	1	TIDAK	USIA < 21	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
15	BERESIKO	1	0	BERESIKO	USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK
16	TIDAK	0	1	TIDAK	USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK

Gambar 5 Hasil Prediksi Algoritma K-Nearest Neighbor

3) Pengujian Data Menggunakan Algoritma Naive Baiyes

Tahapan berikutnya kemudian data diuji menggunakan algoritma Naive Baiyes untuk melihat tingkat akurasi prediksi algoritma Naive Baiyes tersebut dengan memasukan data training untuk melatih algoritma Naive Baiyes dan data testing untuk menguji data dengan algoritma Naive Baiyes



Gambar 6 Pengujian data menggunakan algoritma Naive Baiyes

Setelah data diuji didapat hasil akurasi sebagai berikut:

accuracy: 85.71%			
	true TIDAK	true BERESIKO	class precision
pred. TIDAK	70	0	100.00%
pred. BERESIKO	15	20	57.14%
class recall	82.35%	100.00%	

Gambar 7 Hasil Pengujian menggunakan algoritma Naive Baiyes

Dari hasil pengujian yang dilakukan oleh algoritma Naive Baiyes didapatkan tingkat akurasi sebesar 85.71% dengan hasil prediksi menggunakan Naive Baiyes sebagai berikut :

Row No.	RESIKO	confidence ₁	confidence ₂	prediction ₁	USIA	MAKAWAL	OLAH RAGA	MEKKAH	MEROKOK	SEKSUAL	KELUARGA
1	TIDAK	0.450	0.540	TIDAK	USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
2	BERESIKO	0.741	0.259	BERESIKO	USIA < 21	YA	YA	TIDAK	TIDAK	TIDAK	YA
3	BERESIKO	0.968	0.032	BERESIKO	USIA < 21	YA	YA	TIDAK	TIDAK	YA	TIDAK
4	TIDAK	0.233	0.767	TIDAK	USIA < 17	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK
5	TIDAK	0.424	0.576	TIDAK	USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
6	TIDAK	0.424	0.576	TIDAK	USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
7	TIDAK	0.233	0.767	TIDAK	USIA < 17	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
8	BERESIKO	0.818	0.182	BERESIKO	USIA < 17	YA	YA	TIDAK	YA	YA	TIDAK
9	TIDAK	0.233	0.767	TIDAK	USIA < 17	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
10	TIDAK	0.424	0.576	TIDAK	USIA < 17	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
11	TIDAK	0.166	0.834	TIDAK	USIA < 17	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK
12	TIDAK	0.076	0.924	TIDAK	USIA < 17	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK
13	TIDAK	0.233	0.767	TIDAK	USIA < 17	YA	TIDAK	TIDAK	YA	TIDAK	TIDAK
14	TIDAK	0.358	0.642	TIDAK	USIA < 21	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK
15	BERESIKO	0.925	0.075	BERESIKO	USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA
16	TIDAK	0.460	0.540	TIDAK	USIA < 21	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK

Gambar 8 Hasil Prediksi Algoritma Naïve Bayes

Setelah data diuji dengan kedua algoritma maka dapat dilihat perbandingan tingkat akurasi pengujian kedua algoritma sebagai berikut :

Tabel 1 Perbandingan Hasil Pengujian K-NN dan Naïve Bayes

Algoritma	Accurasy
K-Nearest Neighbor	80.95%
Naïve Bayes	85.71%

Dari kedua hasil pengujian data tersebut dapat dilihat tingkat keakuratan pengujian lebih tinggi pada algoritma Naïve Bayes sebesar 85.71%.

KESIMPULAN

Berdasarkan hasil penelitian di atas, maka dapat diambil kesimpulan sebagai berikut :

1. Tingkat resiko kanker serviks pada remaja di SMKN 1 Banjarmasin masih sangat kecil, dilihat dari hasil kuisioner yang telah dibagikan pada siswi SMKN 1 Banjarmasin dan diuji menggunakan kedua algoritma.
2. Berdasarkan hasil implementasi dan pengukuran algoritma yang diusulkan, diperoleh algoritma terbaik yaitu Naïve Bayes dengan tingkat akurasi 85.71. tetapi akurasi ini masih belum bisa dinilai excellent (sangat baik).

SARAN

Walaupun penelitian ini telah menghasilkan hasil prediksi resiko kanker serviks pada remaja, namun masih sangat perlu dilakukan pengembangan analisis dan hasil lebih lanjut, khususnya memperdalam faktor resiko kanker serviks pada remaja sehingga didapatkan hasil akurasi yang lebih optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan ribuan terima kasih yang tak terhingga kepada Kemen-

ristekdikti yang telah memberikan kesempatan mendapatkan hibah Penelitian Dosen Pemula. Terima kasih pula diucapkan kepada Universitas Islam Kalimantan Muhammad Arsyad Al Banjari yang telah memfasilitasi demi terselesaikannya penelitian ini, Yayasan Kanker Indonesia dan SMKN 1 Banjarmasin yang mendukung penelitian serta terima kasih kepada seluruh rekan-rekan dan semua pihak yang terlibat hingga penulis dapat menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] Aprilia Dian Kartika, dkk. (2013). *Tingkat Pengetahuan Ibu Pendidikan Kesejahteraan Keluarga tentang Kanker Serviks di Badran Kecamatan Jebres Surakarta*. Maternal : Karanganyar.
- [2] Bustami. (128-146). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI: Jurnal Penelitian Teknik Informatika*.
- [3] Imam Rasidi. (2009). *Epidemiologi Kanker Serviks*. Indonesian Journal of Cancer. Tangerang, Banten.
- [4] Lestari, Mei. (2014) Penerapan Algoritma Klasifikasi Nearest Neighbor (K-Nn) Untuk Mendeteksi Penyakit Jantung. Faktor Exacta 7 (4).
- [5] Mei Lestari. (2014). *Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) Untuk Mendeteksi Penyakit Jantung*. Faktor Exacta : Jakarta.
- [6] Musa, Olha & Alang (2017). *Analisis Penyakit Paru-Paru Menggunakan Algoritma K-Nearest Neighbor pada Rumah Sakit Aloe Saboe Kota Gorontalo*. ILKOM.
- [7] Nugroho, Aris & Subanar. (2013). *Naïve Bayes Classification for Due Date Prediction of Pregnancy Data*. Berkala MIPA.
- [8] Kementerian Kesehatan RI (2015). *Situasi Penyakit Kanker*. Jakarta
- [9] Purnama, Parida & C. Supriyanto (2013). *Deteksi Penyakit Diabetes Type II Dengan Naive Bayes Berbasis Particle Swarm Optimization*. Jurnal Teknologi Informasi Volume 9 Nomor 2.
- [10] Retno Tri Vlandari. 2017. *Data Mining Teori dan Aplikasi Rapidminer*. Gava Media : Yogyakarta.
- [11] Rasjidi, Imam (2009). *Epidemiologi Kanker Serviks*. Indonesian Journal of Cancer Vol. III, No. 3

- [12] Sabransyah, M., dkk (2017) *Aplikasi Metode Naïve Bayes dalam Prediksi Risiko Penyakit Jantung*. Jurnal Eksponensial.
- [13] Silvia Ikmalia Fernanda, dkk. (2017). *Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K-Nearest Neighbor (MKNN)*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.
- [14] SL, T., Ip, W., & Tsang, A. H. (2011). Is Naive Bayes a Good Classifier for Document Classification. *International Journal of Software Engineering and Its Applications* 5.3, 37-46.
- [15] Suyanto, Dr. 2017. *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Informatika : Bandung.